Published in final edited form as:

DNA Repair (Amst). 2015 December ; 36: 146–155. doi:10.1016/j.dnarep.2015.09.018.

Genomic Approaches to DNA repair and Mutagenesis

John J. Wyrick^{a,b} and Steven A. Roberts^a

^aSchool of Molecular Biosciences, Washington State University, Pullman, WA 99164, USA

^bCenter for Reproductive Biology, Washington State University, Pullman, WA 99164, USA

Abstract

DNA damage is a constant threat to cells, causing cytotoxicity as well as inducing genetic alterations. The steady-state abundance of DNA lesions in a cell is minimized by a variety of DNA repair mechanisms, including DNA strand break repair, mismatch repair, nucleotide excision repair, base excision repair, and ribonucleotide excision repair. The efficiencies and mechanisms by which these pathways remove damage from chromosomes have been primarily characterized by investigating the processing of lesions at defined genomic loci, among bulk genomic DNA, on episomal DNA constructs, or using *in vitro* substrates. However, the structure of a chromosome is heterogeneous, consisting of heavily protein-bound heterochromatic regions, open regulatory regions, actively transcribed genes, and even areas of transient single stranded DNA.

Consequently, DNA repair pathways function in a much more diverse set of chromosomal contexts than can be readily assessed using previous methods. Recent efforts to develop whole genome maps of DNA damage, repair processes, and even mutations promise to greatly expand our understanding of DNA repair and mutagenesis. Here we review the current efforts to utilize whole genome maps of DNA damage and mutation to understand how different chromosomal contexts affect DNA excision repair pathways.

Keywords

excision repair; mutagenesis; sequencing; mutation signature

1. INTRODUCTION

Our complex genome is under continual assault by a plethora of DNA damaging agents. Cellular DNA repair pathways are a critical first line of defense against cytotoxic and mutagenic DNA lesions. In order to maintain genome integrity, the DNA repair machinery must not only be able to efficiently and accurately repair these lesions, but also do so in an extraordinary variety of different genomic contexts, ranging from DNA lesions buried

Corresponding Authors: John J. Wyrick (jwyrick@vetmed.wsu.edu) and Steven A. Roberts (sroberts@vetmed.wsu.edu).

CONFLICTS OF INTEREST

J.J.W. is an inventor of patents on ChIP-chip technology, which are licensed to Agilent Technologies.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

within highly compacted heterochromatin to lesions exposed on the template strand of an actively transcribed gene. Inaccurate or inefficient repair in such contexts can give rise to mutations that contribute to the development of cancer and other human diseases [1].

Past studies have characterized repair efficiency by measuring the overall removal of DNA lesions in cells or utilized select genes, genomic loci, or *in vitro* substrates to measure repair activity [2]. These methods have been extremely valuable for deciphering repair pathways and activities within cells. However, measuring repair of bulk DNA or repair at selected loci does not fully capture the range of genomic and chromatin contexts that the repair machinery must grapple with *in vivo*. Similarly, while past mutagenesis studies have primarily used select model genes or damaged substrates to measure mutation rates in cells, considerable additional information can be gleaned through genome-wide surveys of mutation frequency.

Recent advances in genomics technology, including the use of microarrays and high throughput sequencing, have opened new opportunities for genome-wide surveys of DNA damage levels, repair efficiency, and mutation rates. The first half of this review will focus on results from genome-wide studies that have investigated how genomic and chromatin context contributes to DNA lesion formation and subsequent excision repair. The second half of the review will describe genomic approaches for studying mutagenesis frequency. Throughout the review, we will focus on DNA lesions targeted by excision repair pathways.

2. Mapping DNA damage using lesion-specific antibodies

A common method to measure damage levels in bulk genomic DNA is to employ lesion-specific antibodies in an enzyme-linked immunosorbent assay (ELISA) or slot blot assay [3]. A number of lesion-specific antibodies have been developed for such assays. These include monoclonal antibodies recognizing UV-induced DNA damage, such as cyclobutane pyrimidine dimers [4–7], 6-4 photoproducts [7–9], and Dewar isomers [10]; monoclonal antibodies that recognize specific classes of base damage, such as 8-oxoguanine or thymine glycol [11–15]; and monoclonal antibodies recognizing acetylaminofluorene-DNA (AAF-1) adducts [16–18] and cisplatin-DNA crosslinks [19, 20].

Recently, these antibodies have been employed to map the genome-wide distribution of DNA lesions using a variant of the chromatin immunoprecipitation-microarray (ChIP-chip) method. ChIP-chip has been widely used to map genomic binding sites of transcription factors and other DNA-binding proteins in a variety of different species [21, 22]. A number of groups have adapted this methodology to map the genomic distribution of DNA lesions and damage-associated histone modifications [23–25].

2.1. Mapping UV damage frequency and repair in *Saccharomyces cerevisiae*

The Waters and Reed laboratories were the first to adapt this technique to map the genomic distribution of DNA lesions. In their initial study [23], they UV irradiated yeast and isolated genomic DNA either immediately following UV exposure (0 hour) or after allowing the yeast cells to repair for two hours in the dark. As depicted in Figure 1A, the sonicated genomic DNA was immunoprecipitated (IP) with an anti-cyclobutane thymine dimer

antibody, which significantly enriched for UV-damaged DNA (>50-fold enrichment). The IP and input DNA samples were amplified, labeled, and hybridized to Agilent yeast tiling microarrays, which tiled the yeast genome with an average probe-to-probe spacing of 290 nucleotides.

Using this approach, a map was generated showing the initial density of thymine dimer lesions across the yeast genome and the lesion distribution following two hours of repair. There were considerable variations in the initial lesion density across the yeast genome. These variations in ChIP-chip signal were largely predicted by a simple model of cyclobutane pyrimidine dimer (CPD) lesion frequency, which calculated the weighted average of the CPD formation frequencies of nearby pyrimidine dinucleotide sequences. Moreover, by measuring CPD damage levels following two hours of repair, the authors were able to make rough estimates of the NER efficiency throughout the yeast genome [23]. Two lines of evidence validated their approach: first, little if any repair was observed for CPD lesions present in mitochondria, which lack NER; and second, deletion of the *RAD4* gene, which is required for NER, eliminated CPD removal in the nuclear genome. The authors observed wide variations in CPD repair rates across the yeast genome, yet the underlying cause(s) of this variation are unclear.

2.2. Mapping UV damage frequency and repair in the human genome

A similar method has been used to map UV-induced CPD lesions in the human genome [24]. In this study, Zavala et al. used tiling microarrays to map CPD damage along human chromosomes 1 and 6, comprising ~400 Mbp of genomic sequence. These microarrays had a considerably higher density of probes than the yeast study, with an average probe-to-probe spacing of 35 bp. This high resolution enabled the authors to identify and characterize CPD hotspots, which were defined as regions with high levels of CPD lesions (>3-fold enrichment relative to input and to IP signal from undamaged cells).

The authors mapped CPD lesions in immortalized human fibroblasts immediately following exposure to 12 J/m² dose of UV radiation. This analysis identified 1128 and 611 CPD hotspots on chromosome 1 and 6, respectively [24]. Intriguingly, the authors found that the distribution of CPD hotspots is not random, but instead is significantly associated with certain genomic features. For example, a high frequency of CPD hotspots are found among the histone gene clusters on chromosomes 1 (1q21) and 6 (6p21). The strongest association, however, was observed with short interspersed elements (SINE). These repetitive elements are an abundant feature of the human genome [26], but probes representing SINE sequences were not included on the microarray (likely to avoid artifacts due to cross-hybridization). Intriguingly, ~70% of the CPD hotspots abutted an Alu SINE, a frequency more than 10-fold higher than expected by random chance [24]. Because the CPD hotspots were derived from microarray probes covering unique sequences flanking each Alu element, this association is unlikely to be a microarray hybridization artifact due to the high copy number of repeat sequences. Instead, it appears that CPD hotspots preferentially form at many Alu SINEs in the human genome.

Alu SINEs possess a number of distinguishing sequence features, including frequently a long polyA tract that is thought to be a remnant of the process of retrotransposition [26]. The

long polyT tract on the complementary strand should be a high-frequency target for CPD lesions. Indeed, the authors found that many CPD hotspots were immediately adjacent to polyT tracts within Alu elements, suggesting a potential mechanism for CPD hotspot formation. Long polyA:T tracts also disfavor nucleosome formation [27–29], and thus damage in such regions could be more easily recognized and repaired by the NER machinery. It is interesting to speculate that polyA:T tracts in Alu elements and other repeat sequences could potentially function as a genomic sink for UV damage, in a sequence context in which the damage could be readily recognized and repaired.

A separate study [25] used a similar approach to map CPD lesions for a small region of the human genome (10 Mbp of human chromosome 17). It was not reported whether high levels of CPD lesions were associated with Alu repeats. Importantly, this study also mapped the locations of cisplatin (or oxaliplatin) DNA adducts, thus demonstrating that the immunoprecipitation and microarray-based method can be used to map other classes of DNA lesions.

2.3. Modulation of chromatin structure during DNA repair

Although these studies did not have sufficient resolution to characterize how nucleosome context affected damage frequency, they did observe chromatin changes associated with subsequent repair of these lesions. Zavala et al. used micrococcal nuclease (MNase) digestion coupled to tiling microarray analysis to characterize changes in nucleosome positioning and occupancy in human cells following 2 hours of repair after UV radiation [24]. They found a small but significant decrease in nucleosome occupancy associated with repair of CPD ‘hotspots’. This finding is in agreement with past studies, which have shown that chromatin remodeling is associated with, and required for, nucleotide excision repair of lesions within nucleosomes [30–34].

Increased histone acetylation has also been associated with repair of CPD lesions in chromatin [31, 35]. The study by the Waters and Reed laboratories used ChIP-chip to map changes in histone H3 K14 acetylation in yeast during repair of UV- or cisplatin-induced DNA damage [25]. Surprisingly, they observed a very similar overall trend in histone acetylation in the UV and cisplatin damaged cells. In contrast, the distributions of CPDs or cisplatin DNA lesions are quite distinct. It is not clear why distinct DNA lesion profiles yielded unexpectedly similar histone acetylation profiles. Future research is needed to address this question.

2.4. Assessment of current microarray-based methods for DNA damage mapping

Immunoprecipitation of DNA lesions followed by microarray analysis has proven to be a reliable method for mapping damage throughout the genome, as long as an appropriate lesion-specific antibody is available. Its primary limitation is low resolution in resolving the location of the lesions. Identified peaks are typically wide, spanning hundreds of base pairs of genomic sequence, even when using high density microarrays. This is largely due to the size of the sonicated DNA fragments, which typically are at least 200–300 bp in length. Resolution of DNA lesions could be improved by using high-throughput sequencing instead of microarray analysis to analyze the immunoprecipitated DNA. Even higher resolution

could potentially be obtained by performing exonuclease digestion of the immunoprecipitated antibody-DNA complex prior to sequencing, similar to the 'ChIP-exo' technique used for precise protein-DNA interaction mapping [36].

3. Towards DNA damage mapping at base pair resolution

The ultimate goal is to achieve base pair resolution in locating DNA lesions throughout the genome. This level of resolution would not only precisely define the locations of lesions within the genomic sequence, but also enable subsequent bioinformatics analyses that are not possible with low-resolution methods. This could include relating damage frequency or repair efficiency to the rotational or translational positioning of lesions within nucleosomes. Recent reports using methods to generate precise strand breaks at DNA lesions coupled with high throughput sequencing have begun to make progress towards this goal, as described in more detail below.

3.1. Precise DNA damage mapping using repair proteins

During the process of excision repair, repair enzymes generate precise single-stranded DNA (ssDNA) nicks at sites of DNA damage. The Hesselberth laboratory used this feature of excision repair as a method to precisely map the locations of uracil lesions within the *S. cerevisiae* and *E. coli* genomes [37]. Uracil lesions can arise from spontaneous (or enzymatically-catalyzed) cytosine deamination or through misincorporation of deoxyuracil nucleotides during DNA replication. Uracil lesions in the genome are recognized by an uracil DNA glycosylase enzyme, which initiates the base excision repair (BER) pathway by cleaving the N-glycosidic bond connecting the uracil base (reviewed in [38]). The Hesselberth laboratory used purified BER enzymes (i.e., uracil DNA glycosylase (UDG) and endonuclease IV) to generate ssDNA nicks at sites of uracil lesions within isolated genomic DNA as a means for mapping these lesions.

With sufficiently high levels of damage, ssDNA nicks on opposite strands will occur in close enough proximity to yield a double strand break. If such double strand breaks occur at high frequency, they will yield small DNA fragments that are amenable to high-throughput sequencing. The ends of the sequenced fragments would then correspond with the sites of the initial uracil lesions, and thus enable base pair resolution mapping of this DNA lesion (Figure 1B). Hesselberth and colleagues called this experimental strategy *pre-digestion* Excision-seq [37].

For this technique to work successfully, very high levels of DNA damage are needed. In their study [37], Hesselberth and colleagues used *E. coli* or yeast cells that accumulate high-levels of uracil lesions due to a mutation in the uracil DNA glycosylase gene (*udg* or *UNG1*). Furthermore, they greatly stimulated uracil incorporation during replication by either mutating the *dut* (*E. coli*) or *DUT1* (yeast) gene encoding the deoxyuridine triphosphate nucleotidohydrolase (dUTPase) enzyme, which cleanses the cellular dNTP pool of dUTP [39], or by treating the cells with the nucleotide analog 5-fluorouracil, which inhibits thymidylate synthase, thus significantly elevating cellular levels of dUTP and its resulting incorporation into genomic DNA.

Using this method, they mapped millions of sequence reads derived from DNA cleavage at uracil lesions in either the yeast or *E. coli* genomes. The sequencing results indicated high specificity for uracil lesions, as the base immediately upstream of the mapped reads was a 'T' (i.e., a potential site for uracil incorporation, see Figure 1B) in more than 98% of the reads. Interestingly, early replicating sequences in both yeast and *E. coli* tended to have lower levels of uracil incorporation, perhaps representing fluctuations in the dUTP fraction of the dNTP pools during DNA replication [37].

Hesselberth and colleagues also mapped UV damage in the yeast genome using a related strategy, except the UV damage endonuclease (UVDE/Uve1) enzyme from *S. pombe* [40] was used to generate ssDNA nicks at CPDs or 6-4 photoproducts (6-4PPs). Again, very high levels of damage were generated by irradiating yeast cells with a UV dose of 10,000 J/m². This technique was not as specific for CPDs or 6-4PPs, as up to 15% of the mapped reads did not have the expected 5' pyrimidine dimer sequence [37]. This could be due to the substrate specificity of the UVDE enzyme, which has been reported to cleave AP sites as well as some forms of base damage in addition to CPDs and 6-4PPs [41, 42] (see elsewhere in this issue).

3.2. Genome-wide mapping of ribonucleotide incorporation

Ribonucleotide lesions in genomic DNA can occur from the infrequent incorporation of rNTPs during replication or can result from hydroxyl radical damage of the 2' position of the DNA deoxyribose moiety. Ribonucleotide lesions are repaired by a ribonucleotide excision repair (RER) pathway involving the enzyme RNase-H2 [43]. Recently, a number of innovative methods have been developed to map genomic ribonucleotide lesions at base pair resolution and with high sensitivity [44–47]. These methods vary widely in their chosen experimental strategy. Three of the methods (i.e., ribose-seq [46], HydEn-seq [44], and Pu-seq [45]) utilize the sensitivity of ribonucleotides to alkaline conditions to generate ssDNA nicks specifically at ribonucleotide incorporation sites. The resulting ssDNA nicks are subsequently processed for PCR amplification and high-throughput sequencing using distinct methods. Ribose-seq [46] exploits the unique 2',3'-phosphate moiety, which is generated at the 3' end of the strand break by alkaline cleavage, as a substrate for a novel *Arabidopsis* tRNA ligase to generate a ssDNA circle. The circular DNA is protected from subsequent exonuclease digestion, while the unligated background DNA is degraded. HydEn-seq [44] detects the presence of the free 5' OH group generated by alkaline cleavage. The cleaved DNA is denatured, 5' phosphorylated, and ligated to a single-stranded DNA oligonucleotide using T4 RNA ligase. Because HydEn-seq recognizes free 5' ends, there is considerable background of sequence reads initiating from the 5' ends of linear chromosomes [44]. Pu-seq [45] converts the ssDNA fragments to double-stranded DNA by random hexamer synthesis using dUTP instead of dTTP nucleotides. The subsequent double-stranded DNA is blunt end ligated to double-stranded oligonucleotides, and the uracil-containing, newly synthesized strand is degraded. The fourth method, embedded ribonucleotide sequencing (emRiboSeq), digests the DNA with purified RNase-H enzyme to generate ssDNA breaks at sites of ribonucleotide incorporation [47]. The resulting ssDNA nicks are processed by ligation of the free 3' OH to a double-stranded oligonucleotide with a

degenerate six-base pair single-stranded overhang and a biotin tag. The resulting ligation products are captured using streptavidin beads and sequenced.

Genome-wide ribonucleotide mapping was primarily used to measure DNA polymerase usage throughout the budding and fission yeast genomes [44–47]. Mutations were generated in replicative polymerases (Pol α , Pol δ , or Pol ϵ) that significantly boosted their ribonucleotide incorporation rate during DNA synthesis. For example, the *pol2-M644G* mutation in Pol ϵ increases its frequency of ribonucleotide incorporation errors by more than 10-fold *in vitro* [48]. Subsequent mapping of ribonucleotide incorporation in a yeast strain background defective in RER (due to a mutant in RNase H2) enabled the investigators to measure which DNA strands were synthesized by each polymerase, using ribonucleotide incorporation as a molecular signature of polymerase usage. These studies confirmed that Pol δ and Pol ϵ were the primary lagging and leading strand polymerases, respectively [44–47].

Interestingly, one study [47] found a significant fraction of Pol α -replicated sequence was retained (~1.5% of the genome), even in non-replicating yeast cells in stationary phase. Normally during lagging strand synthesis the initial DNA primer synthesized by Pol α is displaced by Pol δ as it extends the adjacent Okazaki fragment [49, 50], and is subsequently removed by a flap endonuclease (e.g., FEN1/Rad27 [51]). However, Taylor and colleagues [47] observed that Pol α -synthesized sequence tend to be retained near nucleosome dyad regions and adjacent to strong transcription factor binding sites (e.g., Reb1). They suggested that nucleosome assembly or transcription factor binding could block ongoing Pol δ replication and strand displacement, leading to retention of Pol α -synthesized sequence. Since Pol α is considerably more error-prone than the other replicative polymerases (because it lacks a 3'-5' proofreading exonuclease activity [52]), this could explain the increased mutation rate observed adjacent to nucleosome dyad regions [53–56].

3.3. Assessment of current base pair resolution methods for DNA damage mapping

Pre-digestion excision-seq is the first published genome-wide method able to accurately detect multiple types of DNA damage at base pair resolution [37]. However, this method has a number of drawbacks that may limit its utility going forward. First, very high levels of DNA damage are required to generate sufficient single strand breaks to release short DNA fragments. Hence, it likely cannot be used to map DNA damage occurring at physiological levels in cells. Hasselbreth and colleagues also described a *post-digestion* excision-seq method, which can measure much lower damage levels [37]; however, this method has much lower resolution, on par with the immunoprecipitation-microarray based methods discussed above, and thus will not be discussed further.

Second, pre-digestion excision-seq would be difficult to use for repair rate measurements. This in part due to the high damage levels required for damage detection. For example, CPD repair experiments in human cells typically use a low dose of UV radiation (e.g., 15 J/m²) in order to avoid excessive cell death. In contrast, the excision-seq study used a dose of 10,000 J/m². A more significant concern, however, is that lesion detection requires multiple lesions (i.e., at least four) occurring in close proximity to one another to generate a sequence-able short DNA fragment. Hence, the detection frequency of a lesion depends on the frequency

of adjacent lesions. This is problematic for repair studies. The disappearance of a lesion would not necessarily represent its own repair, but instead could be due to repair of neighboring lesions that are required for its detection.

An ideal damage mapping technique would be sensitive enough to detect single lesions with base pair resolution, as is the case for the ribonucleotide mapping approaches described above [44–47]. Most of these methods exploit the labile nature of RNA nucleotides under alkaline conditions, and thus are not generally applicable to other forms of damage. In the future, perhaps modifications of the excision-seq method or alternative methods could be used to map physiological levels of various types of DNA damage, and its subsequent repair, at base pair resolution.

4. Using mutations to understand how DNA excision repair occurs across the genome

In lieu of single nucleotide resolution mapping of DNA damage, recent efforts to understand how DNA repair processes occur across the genome have turned to utilizing mutation patterns as a means to infer the precise location where a specific type of DNA lesion has occurred. DNA damage induced mutations result from either the lack of DNA repair acting on the lesion or an error-prone repair attempt. Consequently, mutations are a direct biological output of DNA damage and repair events that oftentimes translate to the very same processes underlying genetic alterations that cause diseases like cancer. The contextual analysis of damage induced mutations has proven to be a useful method to gain insight into lesion formation and repair dynamics on a whole-genome scale in both model systems and clinical cancer samples.

4.1. Mutation signatures associated with DNA damage and excision repair

Inferring DNA damage and repair dynamics from whole-genome mutation data requires prior knowledge of defined mutation patterns (also referred to as “mutation signatures” [57, 58]) for specific DNA damaging agents. Mutation signatures are generally composed of the local nucleotide sequences prone to the formation of mutagenic DNA lesions as well as the type of base substitutions that arise from error-prone processing of the lesion.

The mutation signatures associated with classical DNA damaging agents like UV light or benzo[a]pyrene in tobacco smoke have been accrued through studies that evaluated the nucleotide changes occurring in plasmid DNAs containing site-specific DNA lesions or in mutation reporter genes following the exposure of model organisms or cell lines to a given mutagen [59]. Such experiments have delineated the sequence preferences as to where adducts are likely to form, the roles of NER in removing these lesions as well as the mutagenic and error-free bypass of lesions by damage tolerance mechanisms, which all combine to produce a mutation signature. In the case of UV light, the combination of the sequence specificity of cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts [60, 61], the error-free bypass of T-T CPDs by DNA pol eta [62–64], and the possible elevated deamination of adducted cytidines [65–67] result in most UV-induced mutations being single and dinucleotide C to T substitutions at CC and TC sequences (Figure 2). Similarly,

the bulky oxidative DNA damage induced by tobacco smoke produces a large overabundance of C to A substitutions [68, 69]. Comparison of these mutation signatures with mutations identified in tumor suppressor genes like *TP53* in human cancers provided the initial evidence for the causative roles of sunlight and tobacco smoke in skin and lung cancers, respectively [59, 68, 70]. Large-scale whole genome and exome sequencing of human tumors has likewise determined that these mutation signatures are pervasive across many melanoma [57, 71–82] and lung cancer genomes [82–88], suggesting that exposure to these DNA damaging agents are the primary cause of mutation in these tumors.

Further sequencing of additional human tumor types has identified mutational patterns consistent with other DNA damage processes like CpG deamination, enzymatic cytidine deamination by both Activation Induced Cytidine Deaminase and APOBEC (apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like) family cytidine deaminases [83, 89, 90], reactive oxygen species (ROS) [91], aristolochic acid [92–94], and even chemotherapeutics like temozolomide [83] (Table 1). Using non-negative matrix factorization to mathematically classify mutations into related groups, Alexandrov et al., identified 21 distinct mutation signatures occurring among nearly 7000 tumor samples [83]. Since detailed knowledge of the mutation spectra induced by DNA damaging agents is available for only a limited number of agents, many of the causes of these mutation signatures are still unknown. Consequently, recent efforts to increase the utility of whole genome mutation analysis in understanding the causes of mutations that contribute to carcinogenesis have focused on increasing the number of mutagenic signatures associated with DNA mutagens as well as deficiencies in DNA repair pathways [95–97].

4.2. Regional differences in DNA excision repair efficiency can influence mutation rates in the genome

In instances where a dominant known mutation signature can be observed, the analysis of mutation distributions has begun to describe how chromosome structure can govern DNA damage and repair processes. Initial efforts to utilize next-generation sequencing technology to identify mutations across the genomes of both model organisms and clinical tumor samples has uncovered regional differences in mutation rate, which has been most commonly explained by differences in DNA repair capacity.

Using the DNA alkylating agent, ethyl methanesulfonate (EMS) as a DNA damaging and mutagenizing agent, Parkhomchuk et al, sequenced multiple clonal *Escherichia coli* isolates [98]. They identified stretches of G to A mutations indicative of the expected mis-pairing of a thymine base with EMS-induced O6-ethyl guanine lesions in the template strand during replication. Interestingly, regions of the *E. coli* genome lacked any mutations, indicating that either these regions had experienced faster repair than the mutagenized regions or more time was available to repair lesions before these regions replicated and mutations were induced.

Likewise, chronic exposure to methyl methanesulfonate can induce localized clusters of mutations in yeast [99]. Mutations within these regions primarily involved the substitution of cytidines in the same DNA strand, a pattern most easily explained by failure of base excision repair (BER) to remove mutagenic N3-methyl cytidines in transient single stranded gaps of DNA. Similar clustered mutations have since been reported in human cancers likely

induced by inadvertent activity of APOBEC cytidine deaminases on ssDNA gaps and subsequent failure to complete BER [89, 90, 99–103].

Beyond localized inhibition of repair due to ssDNA, the mutation rate across a variety of human tumor genomes as well as the human germline is heterogeneous, correlating with different aspects of chromosome structure and function [104] (Figure 3). Gene poor genomic regions that replicate late in the cell cycle have higher mutation densities [82, 105–107]. These mutation densities also display a strong negative correlation with the transcription levels of these areas [82].

While the underlying reasons for these regional differences are currently unknown, initial hypotheses include an increased error rate of replication in later replicating areas [105] or the ability of Transcription Coupled Nucleotide Excision Repair (TCR) to remove mutagenic lesions in transcribed areas [104]. Supporting a possible role for TCR in sculpting regional mutation rates, multiple cancer sequencing studies have noted an enrichment of UV, tobacco, and aristolochic acid-induced mutation signatures on the non-transcribed strand of genes in melanomas, lung, and urinary tract cancers respectively, suggesting that TCR is efficiently removing these lesions on the transcribed strand [83, 86, 92, 94].

Chromatin structure itself has also been implicated in dictating regional mutation densities in cancer [104, 108, 109]. Analyzing nearly 84,000 mutations from tumors of four cancer types, Lehner and colleagues found correlations of increased mutation densities from regions of the genome that had repressive histone modifications indicative of heterochromatic areas, while mutation densities anti-correlated with marks of open chromatin [108]. Again, increased accessibility of DNA repair factors to DNA lesions in open chromatin is a likely underlying cause. DNase I hypersensitivity mapping of regulatory regions of low nucleosome occupancy have re-capitulated these results, with open regulatory regions generally displaying reduced mutation frequencies [110]. Sunyaev and colleagues provided further evidence that reduced repair efficiency in densely chromatinized regions is one factor producing the observed mutational heterogeneity [110]. They analyzed melanoma genomes containing somatic mutations in NER genes separately from genomes with full NER capacity and determined that in NER-deficient tumors mutation densities occurred more uniformly across the genome. A similar result was seen in the sequencing and analysis of UV induced mutations associating with repressive histone modifications across the genomes of tumors from XPC deficient individuals [111].

4.3. Assessing mutational analyses to understand DNA excision repair in a genomic context

While the analysis of mutation signatures in relation to chromosome features has advanced our knowledge of DNA damage and repair transactions on a global scale as well as how these processes control heterogeneous mutation frequencies in tumors, these types of analyses have specific limitations. The key limitation is that the statistical analysis of mutation locations requires a large mutation sample size from which to draw conclusions. However, the number of mutations commonly found in biological and clinical samples is small compared to the genome size. Consequently, mutations are typically grouped into relatively large genome regions, which limits the spatial resolution over which correlations

can be compared. Indeed, correlations identified between mutation frequencies and histone modifications are weaker as the window size over which these comparisons were made was decreased presumably due to the decrease in the number of mutations in each bin [108]. Despite providing the precise base pair and even DNA strand where a DNA lesion was processed, the resolution of mutational analyses are ultimately limited by the number of mutations available for analysis. This limitation can be overcome by increasing the number of samples sequenced; however, in many cases this option is unavailable.

In addition, most DNA damages are not converted into mutations. Mutational analyses therefore focus on a minor number of damage events and could be skewed by differential processing of lesions in a population where mutagenic lesions are processed differently than non-mutagenic ones. Thus, the picture of genome wide DNA damage and repair kinetics we see through mutational analyses may prove very different from that observed from analyzing bulk lesions. When analyzing clinical samples, however, such a skew towards the mutagenic processes may be beneficial as a major purpose of these efforts is to understand what processes are contributing to the disease state. Moreover, within these human clinical samples, mutational analyses remain the only method to assess the occurrence of DNA damage and repair.

5. Conclusions

It is clear that recent technological advances are ushering in a new era of DNA repair and mutagenesis research. We anticipate that these new technologies will provide unprecedented insight into the effects of genomic and chromatin context on DNA lesion formation, repair efficiency, and mutagenesis frequency. To fully realize the power of these new approaches, however, will require significant improvements in the resolution and sensitivity of genome-wide DNA damage mapping methods.

Acknowledgments

We thank Dr. Peng Mao, Amelia Hodges, Rithy Meas, Dr. Levi O'Loughlin, and Alexander Brown for helpful comments and suggestions. J.J.W. is funded by a National Institutes of Health grant (2R01ES002614) from NIEHS. S.A.R. is supported by a Pathway to Independence Award (4R00ES022633-02) from NIEHS.

References

1. Ciccia A, Elledge SJ. The DNA damage response: making it safe to play with knives. *Molecular cell*. 2010; 40:179–204. [PubMed: 20965415]
2. Friedberg, EC.; Walker, GC.; Siede, W.; Wood, RD.; Schultz, RA.; Ellenberger, T. *DNA repair and Mutagenesis*. 2. ASM Press; Washington, D.C: 2006.
3. Kobayashi N, Katsumi S, Imoto K, Nakagawa A, Miyagawa S, Furumura M, Mori T. Quantitation and visualization of ultraviolet-induced DNA damage using specific antibodies: application to pigment cell biology. *Pigment cell research / sponsored by the European Society for Pigment Cell Research and the International Pigment Cell Society*. 2001; 14:94–102.
4. Strickland PT, Boyle JM. Characterisation of two monoclonal antibodies specific for dimerised and non-dimerised adjacent thymidines in single stranded DNA. *Photochemistry and photobiology*. 1981; 34:595–601. [PubMed: 7301937]
5. Roza L, van der Wulp KJ, MacFarlane SJ, Lohman PH, Baan RA. Detection of cyclobutane thymine dimers in DNA of human cells with monoclonal antibodies raised against a thymine dimer-

containing tetranucleotide. Photochemistry and photobiology. 1988; 48:627–633. [PubMed: 3241835]

6. Mizuno T, Matsunaga T, Ihara M, Nikaido O. Establishment of a monoclonal antibody recognizing cyclobutane-type thymine dimers in DNA: a comparative study with 64M-1 antibody specific for (6-4)photoproducts. Mutation research. 1991; 254:175–184. [PubMed: 2002812]
7. Mori T, Nakane M, Hattori T, Matsunaga T, Ihara M, Nikaido O. Simultaneous establishment of monoclonal antibodies specific for either cyclobutane pyrimidine dimer or (6-4)photoproduct from the same mouse immunized with ultraviolet-irradiated DNA. Photochemistry and photobiology. 1991; 54:225–232. [PubMed: 1780359]
8. Mori T, Matsunaga T, Hirose T, Nikaido O. Establishment of a monoclonal antibody recognizing ultraviolet light-induced (6-4) photoproducts. Mutation research. 1988; 194:263–270. [PubMed: 2460765]
9. Matsunaga T, Mori T, Nikaido O. Base sequence specificity of a monoclonal antibody binding to (6-4)photoproducts. Mutation research. 1990; 235:187–194. [PubMed: 1692967]
10. Matsunaga T, Hatakeyama Y, Ohta M, Mori T, Nikaido O. Establishment and characterization of a monoclonal antibody recognizing the Dewar isomers of (6-4)photoproducts. Photochemistry and photobiology. 1993; 57:934–940. [PubMed: 8367534]
11. Bessalov IA, Bond JP, Purmal AA, Wallace SS, Melamede RJ. Fabs specific for 8-oxoguanine: control of DNA binding. Journal of molecular biology. 1999; 293:1085–1095. [PubMed: 10547287]
12. Soultanakis RP, Melamede RJ, Bessalov IA, Wallace SS, Beckman KB, Ames BN, Taatjes DJ, Janssen-Heininger YM. Fluorescence detection of 8-oxoguanine in nuclear and mitochondrial DNA of cultured cells using a recombinant Fab and confocal scanning laser microscopy. Free radical biology & medicine. 2000; 28:987–998. [PubMed: 10802231]
13. Leadon SA, Hanawalt PC. Monoclonal antibody to DNA containing thymine glycol. Mutation research. 1983; 112:191–200. [PubMed: 6888408]
14. Chen BX, Hubbard K, Ide H, Wallace SS, Erlanger BF. Characterization of a monoclonal antibody to thymidine glycol monophosphate. Radiation research. 1990; 124:131–136. [PubMed: 2247592]
15. Greferath R, Nehls P. Monoclonal antibodies to thymidine glycol generated by different immunization techniques. Hybridoma. 1997; 16:189–193. [PubMed: 9145322]
16. Le Guern A, Leng M, Kourilsky P. Monoclonal antibodies to DNA modified by the carcinogen N-acetoxy-N-2 acetylaminofluorene. Developments in biological standardization. 1984; 57:409–414. [PubMed: 6084621]
17. Ball SS, Quaranta V, Shadravan F, Walford RL. An ELISA for detection of DNA-bound carcinogen using a monoclonal antibody to N-acetoxy-2-acetylaminofluorene-modified DNA. Journal of immunological methods. 1987; 98:195–200. [PubMed: 3571986]
18. Iwamoto TA, Kobayashi N, Imoto K, Yamamoto A, Nakamura Y, Yamauchi Y, Okumura H, Tanaka A, Hanaoka F, Shibutani S, Miyagawa S, Mori T. In situ detection of acetylaminofluorene-DNA adducts in human cells using monoclonal antibodies. DNA repair. 2004; 3:1475–1482. [PubMed: 15380103]
19. Sundquist WI, Lippard SJ, Stollar BD. Monoclonal antibodies to DNA modified with cis- or trans-diamminedichloroplatinum(II). Proceedings of the National Academy of Sciences of the United States of America. 1987; 84:8225–8229. [PubMed: 2446320]
20. Tilby MJ, Johnson C, Knox RJ, Cordell J, Roberts JJ, Dean CJ. Sensitive detection of DNA modifications induced by cisplatin and carboplatin in vitro and in vivo using a monoclonal antibody. Cancer research. 1991; 51:123–129. [PubMed: 1703029]
21. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA. Transcriptional regulatory code of a eukaryotic genome. Nature. 2004; 431:99–104. [PubMed: 15343339]
22. Hanlon SE, Lieb JD. Progress and challenges in profiling the dynamics of chromatin and transcription factor binding with DNA microarrays. Current opinion in genetics & development. 2004; 14:697–705. [PubMed: 15531167]

23. Teng Y, Bennett M, Evans KE, Zhuang-Jackson H, Higgs A, Reed SH, Waters R. A novel method for the genome-wide high resolution analysis of DNA damage. *Nucleic acids research*. 2011; 39:e10. [PubMed: 21062813]
24. Zavala AG, Morris RT, Wyrick JJ, Smerdon MJ. High-resolution characterization of CPD hotspot formation in human fibroblasts. *Nucleic acids research*. 2014; 42:893–905. [PubMed: 24137003]
25. Powell JR, Bennett MR, Evans KE, Yu S, Webster RM, Waters R, Skinner N, Reed SH. 3D-DIP-Chip: a microarray-based method to measure genomic DNA damage. *Scientific reports*. 2015; 5:7975. [PubMed: 25609656]
26. Deininger PL, Batzer MA. Mammalian retroelements. *Genome research*. 2002; 12:1455–1465. [PubMed: 12368238]
27. Suter B, Schnappauf G, Thoma F. Poly(dA:dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. *Nucleic acids research*. 2000; 28:4083–4089. [PubMed: 11058103]
28. Barbic A, Zimmer DP, Crothers DM. Structural origins of adenine-tract bending. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100:2369–2373. [PubMed: 12586860]
29. Segal E, Widom J. Poly(dA:dT) tracts: major determinants of nucleosome organization. *Current opinion in structural biology*. 2009; 19:65–71. [PubMed: 19208466]
30. Osley MA, Tsukuda T, Nickoloff JA. ATP-dependent chromatin remodeling factors and DNA damage repair. *Mutation research*. 2007; 618:65–80. [PubMed: 17291544]
31. Waters R, Teng Y, Yu Y, Yu S, Reed SH. Tilting at windmills? The nucleotide excision repair of chromosomal DNA. *DNA repair*. 2009; 8:146–152. [PubMed: 19041427]
32. Palomera-Sanchez Z, Zurita M. Open, repair and close again: chromatin dynamics and the response to UV-induced DNA damage. *DNA repair*. 2011; 10:119–125. [PubMed: 21130713]
33. Czaja W, Mao P, Smerdon MJ. The Emerging Roles of ATP-Dependent Chromatin Remodeling Enzymes in Nucleotide Excision Repair. *International journal of molecular sciences*. 2012; 13:11954–11973. [PubMed: 23109894]
34. Lans H, Marteijs JA, Vermeulen W. ATP-dependent chromatin remodeling in the DNA-damage response. *Epigenetics & chromatin*. 2012; 5:4. [PubMed: 22289628]
35. Nag R, Smerdon MJ. Altering the chromatin landscape for nucleotide excision repair. *Mutation research*. 2009; 682:13–20. [PubMed: 19167517]
36. Rhee HS, Pugh BF. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*. 2011; 147:1408–1419. [PubMed: 22153082]
37. Bryan DS, Ransom M, Adane B, York K, Hesselberth JR. High resolution mapping of modified DNA nucleobases using excision repair enzymes. *Genome research*. 2014; 24:1534–1542. [PubMed: 25015380]
38. Krokan HE, Drablos F, Slupphaug G. Uracil in DNA--occurrence, consequences and repair. *Oncogene*. 2002; 21:8935–8948. [PubMed: 12483510]
39. Vertessy BG, Toth J. Keeping uracil out of DNA: physiological role, structure and catalytic mechanism of dUTPases. *Accounts of chemical research*. 2009; 42:97–106. [PubMed: 18837522]
40. Bowman KK, Sidik K, Smith CA, Taylor JS, Doetsch PW, Freyer GA. A new ATP-independent DNA endonuclease from *Schizosaccharomyces pombe* that recognizes cyclobutane pyrimidine dimers and 6-4 photoproducts. *Nucleic acids research*. 1994; 22:3026–3032. [PubMed: 8065916]
41. Avery AM, Kaur B, Taylor JS, Mello JA, Essigmann JM, Doetsch PW. Substrate specificity of ultraviolet DNA endonuclease (UVDE/Uve1p) from *Schizosaccharomyces pombe*. *Nucleic acids research*. 1999; 27:2256–2264. [PubMed: 10325412]
42. Kanno S, Iwai S, Takao M, Yasui A. Repair of apurinic/apyrimidinic sites by UV damage endonuclease; a repair protein for UV and oxidative damage. *Nucleic acids research*. 1999; 27:3096–3103. [PubMed: 10454605]
43. Williams JS, Kunkel TA. Ribonucleotides in DNA: origins, repair and consequences. *DNA repair*. 2014; 19:27–37. [PubMed: 24794402]
44. Clausen AR, Lujan SA, Burkholder AB, Orebaugh CD, Williams JS, Clausen MF, Malc EP, Mieczkowski PA, Fargo DC, Smith DJ, Kunkel TA. Tracking replication enzymology in vivo by

genome-wide mapping of ribonucleotide incorporation. *Nature structural & molecular biology*. 2015; 22:185–191.

45. Daigaku Y, Keszthelyi A, Muller CA, Miyabe I, Brooks T, Retkute R, Hubank M, Nieduszynski CA, Carr AM. A global profile of replicative polymerase usage. *Nature structural & molecular biology*. 2015; 22:192–198.
46. Koh KD, Balachander S, Hesselberth JR, Storici F. Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA. *Nature methods*. 2015; 12:251–257. [PubMed: 25622106]
47. Reijns MA, Kemp H, Ding J, de Proce SM, Jackson AP, Taylor MS. Lagging-strand replication shapes the mutational landscape of the genome. *Nature*. 2015; 518:502–506. [PubMed: 25624100]
48. Nick McElhinny SA, Kumar D, Clark AB, Watt DL, Watts BE, Lundstrom EB, Johansson E, Chabes A, Kunkel TA. Genome instability due to ribonucleotide incorporation into DNA. *Nature chemical biology*. 2010; 6:774–781. [PubMed: 20729855]
49. Maga G, Villani G, Tillement V, Stucki M, Locatelli GA, Frouin I, Spadari S, Hubscher U. Okazaki fragment processing: modulation of the strand displacement activity of DNA polymerase delta by the concerted action of replication protein A, proliferating cell nuclear antigen, and flap endonuclease-1. *Proceedings of the National Academy of Sciences of the United States of America*. 2001; 98:14298–14303. [PubMed: 11724925]
50. Pavlov YI, Frahm C, Nick McElhinny SA, Niimi A, Suzuki M, Kunkel TA. Evidence that errors made by DNA polymerase alpha are corrected by DNA polymerase delta. *Current biology : CB*. 2006; 16:202–207. [PubMed: 16431373]
51. Zheng L, Shen B. Okazaki fragment maturation: nucleases take centre stage. *Journal of molecular cell biology*. 2011; 3:23–30. [PubMed: 21278448]
52. Kunkel TA. Balancing eukaryotic replication asymmetry with replication fidelity. *Current opinion in chemical biology*. 2011; 15:620–626. [PubMed: 21862387]
53. Warnecke T, Batada NN, Hurst LD. The impact of the nucleosome code on protein-coding sequence evolution in yeast. *PLoS genetics*. 2008; 4:e1000250. [PubMed: 18989456]
54. Washietl S, Machne R, Goldman N. Evolutionary footprints of nucleosome positions in yeast. *Trends in genetics : TIG*. 2008; 24:583–587. [PubMed: 18951646]
55. Sasaki S, Mello CC, Shimada A, Nakatani Y, Hashimoto S, Ogawa M, Matsushima K, Gu SG, Kasahara M, Ahsan B, Sasaki A, Saito T, Suzuki Y, Sugano S, Kohara Y, Takeda H, Fire A, Morishita S. Chromatin-associated periodicity in genetic variation downstream of transcriptional start sites. *Science*. 2009; 323:401–404. [PubMed: 19074313]
56. Prendergast JG, Semple CA. Widespread signatures of recent selection linked to nucleosome positioning in the human lineage. *Genome research*. 2011; 21:1777–1787. [PubMed: 21903742]
57. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. *Cell reports*. 2013; 3:246–259. [PubMed: 23318258]
58. Poon S, McPherson J, Tan P, Teh B, Rozen S. Mutation signatures of carcinogen exposure: genome-wide detection and new opportunities for cancer prevention. *Genome Medicine*. 2014; 6:24. [PubMed: 25031618]
59. Pfeifer GP, You YH, Besaratinia A. Mutations induced by ultraviolet light. *Mutation research*. 2005; 571:19–31. [PubMed: 15748635]
60. Sage E. Distribution and repair of photolesions in DNA: genetic consequences and the role of sequence context. *Photochemistry and photobiology*. 1993; 57:163–174. [PubMed: 8389052]
61. Pfeifer GP. Formation and processing of UV photoproducts: effects of DNA sequence and chromatin environment. *Photochemistry and photobiology*. 1997; 65:270–283. [PubMed: 9066304]
62. Johnson RE, Washington MT, Prakash S, Prakash L. Fidelity of human DNA polymerase eta. *The Journal of biological chemistry*. 2000; 275:7447–7450. [PubMed: 10713043]
63. Masutani C, Kusumoto R, Iwai S, Hanaoka F. Mechanisms of accurate translesion synthesis by human DNA polymerase eta. *The EMBO journal*. 2000; 19:3100–3109. [PubMed: 10856253]
64. Johnson RE, Prakash S, Prakash L. Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Pol eta. *Science*. 1999; 283:1001–1004. [PubMed: 9974380]

65. Tu Y, Dammann R, Pfeifer GP. Sequence and time-dependent deamination of cytosine bases in UVB-induced cyclobutane pyrimidine dimers in vivo. *Journal of molecular biology*. 1998; 284:297–311. [PubMed: 9813119]
66. Burger A, Fix D, Liu H, Hays J, Bockrath R. In vivo deamination of cytosine-containing cyclobutane pyrimidine dimers in *E. coli*: a feasible part of UV-mutagenesis. *Mutation research*. 2003; 522:145–156. [PubMed: 12517420]
67. Peng W, Shaw BR. Accelerated deamination of cytosine residues in UV-induced cyclobutane pyrimidine dimers leads to CC-->TT transitions. *Biochemistry*. 1996; 35:10172–10181. [PubMed: 8756482]
68. Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, Hainaut P. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene*. 2002; 21:7435–7451. [PubMed: 12379884]
69. Yoon JH, Besaratinia A, Feng Z, Tang MS, Amin S, Luch A, Pfeifer GP. DNA damage, repair, and mutation induction by (+)-Syn and (-)-anti-dibenzo[a,l]pyrene-11,12-diol-13,14-epoxides in mouse cells. *Cancer research*. 2004; 64:7321–7328. [PubMed: 15492252]
70. Pfeifer GP, Besaratinia A. Mutational spectra of human cancer. *Human genetics*. 2009; 125:493–506. [PubMed: 19308457]
71. Berger MF, Hodis E, Heffernan TP, Deribe YL, Lawrence MS, Protopopov A, Ivanova E, Watson IR, Nickerson E, Ghosh P, Zhang H, Zeid R, Ren X, Cibulskis K, Sivachenko AY, Wagle N, Sucker A, Sougnez C, Onofrio R, Ambrogio L, Auclair D, Fennell T, Carter SL, Drier Y, Stojanov P, Singer MA, Voet D, Jing R, Saksena G, Barretina J, Ramos AH, Pugh TJ, Stransky N, Parkin M, Winckler W, Mahan S, Ardlie K, Baldwin J, Wargo J, Schadendorf D, Meyerson M, Gabriel SB, Golub TR, Wagner SN, Lander ES, Getz G, Chin L, Garraway LA. Melanoma genome sequencing reveals frequent PREX2 mutations. *Nature*. 2012; 485:502–506. [PubMed: 22622578]
72. Furney SJ, Turajlic S, Stamp G, Nohadani M, Carlisle A, Thomas JM, Hayes A, Strauss D, Gore M, van den Oord J, Larkin J, Marais R. Genome sequencing of mucosal melanomas reveals that they are driven by distinct mechanisms from cutaneous melanoma. *The Journal of pathology*. 2013; 230:261–269. [PubMed: 23620124]
73. Furney SJ, Turajlic S, Stamp G, Thomas JM, Hayes A, Strauss D, Gavrielides M, Xing W, Gore M, Larkin J, Marais R. The mutational burden of acral melanoma revealed by whole-genome sequencing and comparative analysis. *Pigment cell & melanoma research*. 2014; 27:835–838. [PubMed: 24913711]
74. Gartner JJ, Davis S, Wei X, Lin JC, Trivedi NS, Teer JK, Program NCS, Meltzer PS, Rosenberg SA, Samuels Y. Comparative exome sequencing of metastatic lesions provides insights into the mutational progression of melanoma. *BMC genomics*. 2012; 13:505. [PubMed: 23006843]
75. Gartner JJ, Parker SC, Prickett TD, Dutton-Regester K, Stitzel ML, Lin JC, Davis S, Simhadri VL, Jha S, Katagiri N, Gotea V, Teer JK, Wei X, Morken MA, Bhanot UK, Program NCS, Chen G, Elnitski LL, Davies MA, Gershenwald JE, Carter H, Karchin R, Robinson W, Robinson S, Rosenberg SA, Collins FS, Parmigiani G, Komar AA, Kimchi-Sarfaty C, Hayward NK, Margulies EH, Samuels Y. Whole-genome sequencing identifies a recurrent functional synonymous mutation in melanoma. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110:13481–13486. [PubMed: 23901115]
76. Krauthammer M, Kong Y, Ha BH, Evans P, Bacchiocchi A, McCusker JP, Cheng E, Davis MJ, Goh G, Choi M, Ariyan S, Narayan D, Dutton-Regester K, Capatana A, Holman EC, Bosenberg M, Sznol M, Kluger HM, Brash DE, Stern DF, Materin MA, Lo RS, Mane S, Ma S, Kidd KK, Hayward NK, Lifton RP, Schlessinger J, Boggon TJ, Halaban R. Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. *Nature genetics*. 2012; 44:1006–1014. [PubMed: 22842228]
77. Kunz M, Dannemann M, Kelso J. High-throughput sequencing of the melanoma genome. *Experimental dermatology*. 2013; 22:10–17. [PubMed: 23174022]
78. Martin M, Masshofer L, Temming P, Rahmann S, Metz C, Bornfeld N, van de Nes J, Klein-Hitpass L, Hinnebusch AG, Horsthemke B, Lohmann DR, Zeschnigk M. Exome sequencing identifies recurrent somatic mutations in EIF1AX and SF3B1 in uveal melanoma with disomy 3. *Nature genetics*. 2013; 45:933–936. [PubMed: 23793026]

79. Nikolaev SI, Rimoldi D, Iseli C, Valsesia A, Robyr D, Gehrig C, Harshman K, Guipponi M, Bukach O, Zoete V, Michielin O, Muehlethaler K, Speiser D, Beckmann JS, Xenarios I, Halazonetis TD, Jongeneel CV, Stevenson BJ, Antonarakis SE. Exome sequencing identifies recurrent somatic MAP2K1 and MAP2K2 mutations in melanoma. *Nature genetics*. 2012; 44:133–139. [PubMed: 22197931]
80. Stark MS, Woods SL, Gartside MG, Bonazzi VF, Dutton-Regester K, Aoude LG, Chow D, Sereduk C, Niemi NM, Tang N, Ellis JJ, Reid J, Zismann V, Tyagi S, Muzny D, Newsham I, Wu Y, Palmer JM, Pollak T, Youngkin D, Brooks BR, Lanagan C, Schmidt CW, Kobe B, MacKeigan JP, Yin H, Brown KM, Gibbs R, Trent J, Hayward NK. Frequent somatic mutations in MAP3K5 and MAP3K9 in metastatic melanoma identified by exome sequencing. *Nature genetics*. 2012; 44:165–169. [PubMed: 22197930]
81. Wei X, Walia V, Lin JC, Teer JK, Prickett TD, Gartner J, Davis S, Program NCS, Stemke-Hale K, Davies MA, Gershenwald JE, Robinson W, Robinson S, Rosenberg SA, Samuels Y. Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nature genetics*. 2011; 43:442–446. [PubMed: 21499247]
82. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, Drier Y, Zou L, Ramos AH, Pugh TJ, Stransky N, Helman E, Kim J, Sougnez C, Ambrogio L, Nickerson E, Shefler E, Cortes ML, Auclair D, Saksena G, Voet D, Noble M, DiCara D, Lin P, Lichtenstein L, Heiman DI, Fennell T, Imielinski M, Hernandez B, Hodis E, Baca S, Dulak AM, Lohr J, Landau DA, Wu CJ, Melendez-Zajgla J, Hidalgo-Miranda A, Koren A, McCarroll SA, Mora J, Lee RS, Crompton B, Onofrio R, Parkin M, Winckler W, Ardlie K, Gabriel SB, Roberts CW, Biegel JA, Stegmaier K, Bass AJ, Garraway LA, Meyerson M, Golub TR, Gordenin DA, Sunyaev S, Lander ES, Getz G. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013; 499:214–218. [PubMed: 23770567]
83. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjord JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Illic T, Imbeaud S, Imielinski M, Jager N, Jones DT, Jones D, Knappskog S, Kool M, Lakhani SR, Lopez-Otin C, Martin S, Munshi NC, Nakamura H, Northcott PA, Pajic M, Papaemmanuil E, Paradiso A, Pearson JV, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt AN, Valdes-Mas R, van Buuren MM, van 't Veer L, Vincent-Salomon A, Waddell N, Yates LR, Zucman-Rossi J, Futreal PA, McDermott U, Lichter P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ, Stratton MR. Signatures of mutational processes in human cancer. *Nature*. 2013; 500:415–421. [PubMed: 23945592]
84. Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, Maher CA, Fulton R, Fulton L, Wallis J, Chen K, Walker J, McDonald S, Bose R, Ornitz D, Xiong D, You M, Dooling DJ, Watson M, Mardis ER, Wilson RK. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell*. 2012; 150:1121–1134. [PubMed: 22980976]
85. Lee W, Jiang Z, Liu J, Haverty PM, Guan Y, Stinson J, Yue P, Zhang Y, Pant KP, Bhatt D, Ha C, Johnson S, Kennemer MI, Mohan S, Nazarenko I, Watanabe C, Sparks AB, Shames DS, Gentleman R, de Sauvage FJ, Stern H, Pandita A, Ballinger DG, Drmanac R, Modrusan Z, Seshagiri S, Zhang Z. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature*. 2010; 465:473–477. [PubMed: 20505728]
86. Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, Lin ML, Beare D, Lau KW, Greenman C, Varela I, Nik-Zainal S, Davies HR, Odonez GR, Mudie LJ, Latimer C, Edkins S, Stebbings L, Chen L, Jia M, Leroy C, Marshall J, Menzies A, Butler A, Teague JW, Mangion J, Sun YA, McLaughlin SF, Peckham HE, Tsung EF, Costa GL, Lee CC, Minna JD, Gazdar A, Birney E, Rhodes MD, McKernan KJ, Stratton MR, Futreal PA, Campbell PJ. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature*. 2010; 463:184–190. [PubMed: 20016488]
87. Bass AJ, Watanabe H, Mermel CH, Yu S, Perner S, Verhaak RG, Kim SY, Wardwell L, Tamayo P, Gat-Viks I, Ramos AH, Woo MS, Weir BA, Getz G, Beroukhi R, O'Kelly M, Dutt A, Rozenblatt-Rosen O, Dziunycz P, Komisarof J, Chirieac LR, Lafargue CJ, Scheble V, Wilbertz T, Ma C, Rao S, Nakagawa H, Stairs DB, Lin L, Giordano TJ, Wagner P, Minna JD, Gazdar AF, Zhu

- CQ, Brose MS, Cecconello I, UR, Marie SK, Dahl O, Shivdasani RA, Tsao MS, Rubin MA, Wong KK, Regev A, Hahn WC, Beer DG, Rustgi AK, Meyerson M. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nature genetics*. 2009; 41:1238–1242. [PubMed: 19801978]
88. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A, Sougnez C, Auclair D, Lawrence MS, Stojanov P, Cibulskis K, Choi K, de Waal L, Sharifnia T, Brooks A, Greulich H, Banerji S, Zander T, Seidel D, Leenders F, Ansen S, Ludwig C, Engel-Riedel W, Stoelben E, Wolf J, Goparju C, Thompson K, Winckler W, Kwiatkowski D, Johnson BE, Janne PA, Miller VA, Pao W, Travis WD, Pass HI, Gabriel SB, Lander ES, Thomas RK, Garraway LA, Getz G, Meyerson M. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*. 2012; 150:1107–1120. [PubMed: 22980975]
89. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, Kiezun A, Kryukov GV, Carter SL, Saksena G, Harris S, Shah RR, Resnick MA, Getz G, Gordenin DA. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nature genetics*. 2013; 45:970–976. [PubMed: 23852170]
90. Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nature genetics*. 2013
91. Bacolla A, Temiz NA, Yi M, Ivanic J, Cer RZ, Donohue DE, Ball EV, Mudunuri US, Wang G, Jain A, Volfovsky N, Luke BT, Stephens RM, Cooper DN, Collins JR, Vasquez KM. Guanine holes are prominent targets for mutation in cancer and inherited disease. *PLoS genetics*. 2013; 9:e1003816. [PubMed: 24086153]
92. Hoang ML, Chen CH, Sidorenko VS, He J, Dickman KG, Yun BH, Moriya M, Niknafs N, Douville C, Karchin R, Turesky RJ, Pu YS, Vogelstein B, Papadopoulos N, Grollman AP, Kinzler KW, Rosenquist TA. Mutational signature of aristolochic acid exposure as revealed by whole-exome sequencing. *Science translational medicine*. 2013; 5:197ra102.
93. Hollstein M, Moriya M, Grollman AP, Olivier M. Analysis of TP53 mutation spectra reveals the fingerprint of the potent environmental carcinogen, aristolochic acid. *Mutation research*. 2013; 753:41–49. [PubMed: 23422071]
94. Poon SL, Pang ST, McPherson JR, Yu W, Huang KK, Guan P, Weng WH, Siew EY, Liu Y, Heng HL, Chong SC, Gan A, Tay ST, Lim WK, Cutcutache I, Huang D, Ler LD, Nairismagi ML, Lee MH, Chang YH, Yu KJ, Chan-On W, Li BK, Yuan YF, Qian CN, Ng KF, Wu CF, Hsu CL, Bunte RM, Stratton MR, Futreal PA, Sung WK, Chuang CK, Ong CK, Rozen SG, Tan P, Teh BT. Genome-wide mutational signatures of aristolochic acid and its application as a screening tool. *Science translational medicine*. 2013; 5:197ra101.
95. Stirling PC, Shen Y, Corbett R, Jones SJ, Hieter P. Genome destabilizing mutator alleles drive specific mutational trajectories in *Saccharomyces cerevisiae*. *Genetics*. 2014; 196:403–412. [PubMed: 24336748]
96. Meier B, Cooke SL, Weiss J, Bailly AP, Alexandrov LB, Marshall J, Raine K, Maddison M, Anderson E, Stratton MR, Gartner A, Campbell PJ. *C. elegans* whole-genome sequencing reveals mutational signatures related to carcinogens and DNA repair deficiency. *Genome research*. 2014; 24:1624–1636. [PubMed: 25030888]
97. Serero A, Jubin C, Loeillet S, Legoix-Ne P, Nicolas AG. Mutational landscape of yeast mutator strains. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111:1897–1902. [PubMed: 24449905]
98. Parkhomchuk D, Amstislavskiy V, Soldatov A, Ogryzko V. Use of high throughput sequencing to observe genome dynamics at a single cell level. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:20830–20835. [PubMed: 19934054]
99. Roberts SA, Sterling J, Thompson C, Harris S, Mav D, Shah R, Klimczak LJ, Kryukov GV, Male E, Mieczkowski PA, Resnick MA, Gordenin DA. Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Molecular cell*. 2012; 46:424–435. [PubMed: 22607975]
100. Early Breast Cancer Trialists' Collaborative Group. Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. An overview of 61 randomized trials

among 28,896 women. The New England journal of medicine. 1988; 319:1681–1692. [PubMed: 3205265]

101. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, Menzies A, Martin S, Leung K, Chen L, Leroy C, Ramakrishna M, Rance R, Lau KW, Mudie LJ, Varela I, McBride DJ, Bignell GR, Cooke SL, Shlien A, Gamble J, Whitmore I, Maddison M, Tarpey PS, Davies HR, Papaemmanuil E, Stephens PJ, McLaren S, Butler AP, Teague JW, Jonsson G, Garber JE, Silver D, Miron P, Fatima A, Boyault S, Langerod A, Tutt A, Martens JW, Aparicio SA, Borg A, Salomon AV, Thomas G, Borresen-Dale AL, Richardson AL, Neuberger MS, Futreal PA, Campbell PJ, Stratton MR. Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell*. 2012; 149:979–993. [PubMed: 22608084]
102. Taylor BJ, Nik-Zainal S, Wu YL, Stebbings LA, Raine K, Campbell PJ, Rada C, Stratton MR, Neuberger MS. DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *Elife*. 2013; 2:e00534. [PubMed: 23599896]
103. Lada AG, Dhar A, Boissy RJ, Hirano M, Rubel AA, Rogozin IB, Pavlov YI. AID/APOBEC cytosine deaminase induces genome-wide kataegis. *Biology direct*. 2012; 7:47. discussion 47. [PubMed: 23249472]
104. Hodgkinson A, Chen Y, Eyre-Walker A. The large-scale distribution of somatic mutations in cancer genomes. *Human mutation*. 2012; 33:136–143. [PubMed: 21953857]
105. Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov GV, Mirkin SM, Sunyaev SR. Human mutation rate associated with DNA replication timing. *Nature genetics*. 2009; 41:393–395. [PubMed: 19287383]
106. Woo YH, Li WH. DNA replication timing and selection shape the landscape of nucleotide variation in cancer genomes. *Nature communications*. 2012; 3:1004.
107. Koren A, Polak P, Nemesh J, Michaelson JJ, Sebat J, Sunyaev SR, McCarroll SA. Differential relationship of DNA replication timing to different forms of human mutation and variation. *Am J Hum Genet*. 2012; 91:1033–1040. [PubMed: 23176822]
108. Schuster-Bockler B, Lehner B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature*. 2012; 488:504–507. [PubMed: 22820252]
109. Liu L, De S, Michor F. DNA replication timing and higher-order nuclear organization determine single-nucleotide substitution patterns in cancer genomes. *Nature communications*. 2013; 4:1502.
110. Polak P, Lawrence MS, Haugen E, Stoletzki N, Stojanov P, Thurman RE, Garraway LA, Mirkin S, Getz G, Stamatoyannopoulos JA, Sunyaev SR. Reduced local mutation density in regulatory DNA of cancer genomes is linked to DNA repair. *Nature biotechnology*. 2014; 32:71–75.
111. Zheng CL, Wang NJ, Chung J, Moslehi H, Sanborn JZ, Hur JS, Collisson EA, Vemula SS, Naujokas A, Chiotti KE, Cheng JB, Fassihi H, Blumberg AJ, Bailey CV, Fudem GM, Mihm FG, Cunningham BB, Neuhaus IM, Liao W, Oh DH, Cleaver JE, LeBoit PE, Costello JF, Lehmann AR, Gray JW, Spellman PT, Arron ST, Huh N, Purdom E, Cho RJ. Transcription restores DNA repair to heterochromatin, determining regional mutation rates in cancer genomes. *Cell reports*. 2014; 9:1228–1234. [PubMed: 25456125]

Highlights

- The genomic and chromatin contexts of a DNA lesion can impact damage frequency and repair outcomes: lessons from genomics
- Survey of genome-wide methods for mapping different types of DNA lesions in yeast and human cells
- Analysis of damage-associated mutation signatures in cancer can be used to understand how chromatin structure influences repair in a variety of contexts
- Achieving base-pair resolution in genome-wide mapping of DNA lesions: progress and current limitations

PROLOGUE

I (JJW) first met Mick Smerdon when I was an undergraduate student at Washington State University (1992–1996), and was delighted to later become one of his colleagues when I joined WSU as an assistant professor in 2002. Mick is a great resource for a beginning faculty member, and later he was especially instrumental when I transitioned into studying aspects of DNA repair, in collaboration with his team. Mick showed me ‘the ropes’ in this new field, and some of my favorite science memories revolve around joint lab meetings with Mick’s group. These meetings were always informative, and often entertaining. Mick is an outstanding scientist and an ideal colleague. I am very grateful to have had the opportunity to work on such interesting projects with his talented team, and learn so much from him in the process!

I (SAR) met Mick at a conference while I was a graduate student studying the impacts of nucleosome structure on DNA double strand break repair by non-homologous end joining. The work from Mick’s lab has served as inspiration and a guide to assessing how DNA repair processes function in the context of chromatin to countless researchers, myself included. As a young independent investigator studying DNA repair and mutagenesis at WSU, I have likewise benefited from Mick’s collegial nature, his penetrating scientific insight, and his support as a mentor.

Fig. 1. Schematic showing experimental strategies for mapping DNA lesions throughout the genome

A. Immunoprecipitation followed by microarray analysis.

B. Pre-digestion Excision-seq [37].

Cyclobutane pyrimidine dimers (CPD) are represented with a T=T symbol in brown, while 6-4 photoproducts (6-4PP) are in blue. Deoxyuracil lesions (dUracil) are in green. In the Excision-seq method, the damaged DNA is initially cleaved with uracil DNA glycosylase (UDG) and Endonuclease IV (Endo IV). This example shows the expected sequencing results for mapping dUracil lesions arising due to misincorporation during replication

Fig. 2. DNA damage and repair influence mutation signatures

Where DNA damage occurs as well as how it is processed and repaired determine the spectrum of mutations that results from exposure to a DNA damaging agent. For example, UV radiation occurs at specific dinucleotides: TT, TC, CC, and CT (not pictured). Nucleotide Excision Repair (NER) can remove these lesions, but in some contexts, such as when a lesion resides in dense chromatin, repair may be inefficient allowing subsequent processes to convert these lesions into mutations (red letters). Unrepaired T-T dimers are frequently bypassed in an error-free manner by DNA polymerase ϵ , while deamination of cytidines in dimers as well as error-prone bypass results in nearly all UV induced mutations occurring at TC and CC dinucleotides (bottom, all possible trinucleotide DNA sequences are shown with the mutated base capitalized, complementary sequences are combined).

Fig. 3. Profiling cancer mutations against genomic features

A. To identify genomic features that associate with elevated mutation densities, the genomes of multiple excised human tumors are sequenced by next-generation sequencing techniques and a list of somatically acquired mutations (red) is compiled by comparison to a patient's germline genome. To call mutations using next-generation sequencing, the mutant allele must be supported by multiple reads in the sequencing data, resulting in the positions of only sub-clonal mutations being determined.

B. In parallel, the locations of various genomic features (e.g. histone modifications (black), chromatin accessibility (light blue), replication timing (purple)) is determined by shot-gun sequencing DNA ends associated with immune-precipitated proteins or DNA modifications (i.e. ChIP-seq, Repli-seq) or DNA ends generated by nucleolytic cleavage (DNase-seq). Cell lines utilized for these experiments are generally chosen to closely resemble the tissue of origin of the tumor whose mutations are analyzed.

C. Correlations between mutation position and any genomic feature is determined by either comparing mutation densities to sequencing tag densities in a given window of genomic space, or by stratifying regions of the genome by the tag densities of a chromosome feature, and subsequently determining the mutation density with each feature category.

Table 1

Mutation signatures associated with DNA damage in human cancer genomes

Source of mutations	Mutated motifs	Resulting base(s)
UV-radiation	YC; Y <u>C</u> G; <u>C</u> C	T; T; TT
5-meCpG deamination	<u>C</u> G	T
APOBEC	T <u>C</u> W	T or G
AID	WRC <u>C</u>	T or G
ROS	<u>C</u>	A
Tobacco	<u>C</u>	A
Temozolomide	<u>C</u> Y	A
Aristolochic acid	CT <u>R</u>	A

* mutated base is underlined. W = A or T, Y = C or T, R = A or G.